

STUDIA METODOLOGICZNE

Henryk GURGUL
Paweł ZAJĄC

Modelowanie nieparametryczne czasu przeżycia przedsiębiorstw w Małopolsce

Streszczenie. *Celem artykułu jest ocena czasu przeżycia firm działających w województwie małopolskim w latach 2006—2014. W badaniu zastosowano metodę analizy trwania.*

Wyniki badania potwierdziły, że podmioty gospodarki narodowej o większej liczbie pracowników dłużej prowadziły działalność, natomiast likwidacja najczęściej dotyczyła firm ze sfery finansowej i ubezpieczeniowej, a najrzadziej — administracyjnej. W ujęciu przestrzennym działalność przedsiębiorstw trwała średnio najdłużej w Krakowie oraz powiatach ziemskich: krakowskim, wielickim i tatrzańskim. Najkrócej zaś trwały firmy w powiecie nowotarskim, a zakładane w miastach działały średnio dłużej niż na wsiach.

Słowa kluczowe: przedsiębiorstwa, Małopolska, funkcja hazardu, analiza przeżycia.

JEL: C34, D40, G33, L25

Istnieje bezpośrednia zależność pomiędzy liczbą przedsiębiorstw i stopą bezrobocia oraz liczbą miejsc pracy. W warunkach kryzysu gospodarczego zmniejsza się liczba miejsc pracy, a zatem wzrasta bezrobocie. Upadłość przedsiębiorstw występuje jednak nie tylko podczas kryzysu gospodarczego, choć wówczas z reguły się nasila (Hol, 2007). Pojawia się ona od początku funkcjonowania gospodarki rynkowej i jest jej nieodłączną cechą. Pozwala na oczyszczenie rynku z podmiotów gospodarczych, które nie mogą dostosować się do jego wymogów i reguł gospodarki rynkowej bądź ich nie szanują. Analiza przeżywalno-

ści przedsiębiorstwa jest ważna m.in. z punktu widzenia ubezpieczeń. Przeżywalność nowych firm oraz możliwość bankructw i ich ekonomiczne konsekwencje są dyskutowane w literaturze przedmiotu (Gurgul, Zając, Matschke i Matschke, 2014). Główną teorią naukową, do której odnoszą się badania empiryczne jest tzw. twórcza destrukcja Schumpetera.

Wprowadzone przez Bakunina, Nietzschego i Sombarta pojęcie twórczej destrukcji zostało upowszechnione przez Schumpetera (1942), który nadał mu treść ekonomiczną. Zdaniem tego autora powstawanie nowych przedsiębiorstw z nowatorskimi metodami produkcji stanowi główny motor długotrwałego rozwoju ekonomicznego. Schumpeter akceptował fakt, że przedsiębiorstwa powstają często na gruzach firm, które wcześniej dominowały na rynku.

Po upadku starych firm ich pracownicy mogą uczestniczyć w pracach bardziej twórczych i produktywnych przedsiębiorstw. W konsekwencji są wytwarzane lepsze produkty, często mniejszym kosztem. Do głównych czynników wywołujących twórczą destrukcję zalicza się nowości dotyczące: produktów lub rynków, sprzętu, źródeł siły roboczej lub surowców, jak również w zakresie nowych metod: organizacji pracy lub zarządzania, zarządzania zapasami oraz dotyczących transportu, komunikacji, reklamy lub marketingu czy instrumentów finansowych. Schumpeter wskazywał zwłaszcza na silny wzrost znaczenia innowacji jako czynnika podnoszenia konkurencyjności przedsiębiorstw na rynku globalnym, a także na dążenie do ulepszania zarządzania poprzez działalność innowacyjną firm w konkurencyjnym i zmieniającym się otoczeniu gospodarczym.

Celem artykułu jest analiza przeżycia firm działających w województwie małopolskim w latach 2006—2014. Według wiedzy autorów powstała jedna publikacja (Śmiech, 2011) na temat przeżycia przedsiębiorstw małopolskich, w której wykorzystano metody bardziej zaawansowane od tabel przeżycia. Mogło to być spowodowane ograniczonym dostępem do danych. Stosowanie bowiem metod analizy trwania wymaga posiadania danych indywidualnych, które nie są publikowane w biuletynach GUS.

PRZEGLĄD LITERATURY

Wśród czynników, które najczęściej wykorzystywane są jako determinanty przeżywalności przedsiębiorstw wyróżnia się: wiek firmy (Gerosky, 1995), rodzaj działalności (Lin i Huang, 2008), miejsce działalności (Gilbert, McDougall i Audretsch, 2006), wielkość firmy (Agarwal i Audretsch, 2001), czynniki makroekonomiczne (Holmes, Hunt i Stone, 2009) oraz indywidualną charakterystykę firmy.

Badacze zgadzają się, że wielkość firmy na początku działalności wpływa w sposób istotny na szansę jej przetrwania (Audretsch i Mahmoon, 1995). Nunes i Sarmiento (2010) przekonują, że istnieje relacja pomiędzy wielkością firmy w chwili rozpoczynania działalności i spodziewaną wysokością zysków.

W związku z tym mniejsze firmy są likwidowane dużo częściej niż duże. Wśród powodów takiego stanu rzeczy wskazuje się najczęściej na bardziej wykwalifikowaną kadrę zarządzającą w dużych firmach oraz fakt, że posiadają one większy budżet, który pozwala przetrwać okresy kryzysowe. Duże firmy mają do dyspozycji jeszcze jeden sposób, który czasem pozwala uniknąć likwidacji. Jest nim redukcja zatrudnienia. Na niekorzyść małych firm przemawia również ograniczony budżet, mogący skutkować niewystarczającymi inwestycjami.

Pozostaje pytanie, w jaki sposób najlepiej mierzyć wielkość firmy. Mata, Portugal i Guimaraes (1995) twierdzą, że to bieżąca wielkość firmy powinna być wykorzystywana w analizach. Fakt, że firma powiększyła się w nieodległej przeszłości sugeruje, że owa firma znajduje się w dobrej kondycji. Można zatem sądzić, że prawdopodobieństwo jej likwidacji jest małe. Z kolei fakt, że firma jest aktualnie mniejsza niż na początku działalności, może wskazywać na problemy finansowe i większe ryzyko likwidacji. Niektórzy autorzy (López-Garcia i Puente, 2006) sądzą, że zmniejszenie się wielkości firmy jest już pierwszym stadium likwidacji, natomiast wzrost wielkości firmy może być związany z zaciągnięciem kredytu, który w przyszłości może być powodem zakończenia działalności i dlatego właśnie wielkość firmy w chwili jej założenia jest właściwą jej charakterystyką.

Problematyką żywotności przedsiębiorstw zajmowali się również polscy autorzy, którzy wykorzystywali w tym celu metody zarówno jakościowe, jak i ilościowe. We wspomnianym artykule Śmiech (2011) wykorzystał nieparametryczną funkcję przeżycia Kaplana-Meiera i zbadał czas trwania 19647 podmiotów gospodarki narodowej zarejestrowanych w REGON w latach 2002 i 2003. Podmioty te podzielone zostały na grupy w zależności od wielkości (liczba zatrudnionych), formy prawnej i sekcji PKD 2004. Przedmiotem badania była przeżywalność do końca roku 2008, czyli 7–8 lat życia firmy.

Poznańska (2005) opisała uzyskane wspólnie z Jacksonem i Klichem wskaźniki korelacji między liczbą przedsiębiorstw, które przetrwały i rozwinęły się, a wybranymi zmiennymi regionalnymi według województw (np. stopa bezrobocia, liczba oddziałów banków, jakość infrastruktury, gęstość zaludnienia, poziom edukacji).

Zdaniem Mączyńskiej (2009) jedną z najważniejszych barier występujących w gospodarce przemian ukierunkowanych na przeciwdziałanie zaburzeniom i nieprawidłowościom, które w skrajnych przypadkach mogą prowadzić do bankructw przedsiębiorstw i całych sektorów produkcyjnych, jest niedorozwój i niedostosowanie infrastruktury instytucjonalno-regulacyjnej do wymogów nowego wzorca cywilizacyjnego. Jej zdaniem częścią tak rozumianej gospodarki jest instytucjonalna infrastruktura upadłości przedsiębiorstw. Opierając się na wynikach badań empirycznych autorka wskazała na wybrane elementy tej infrastruktury, m.in. takie jak prawo upadłościowe i naprawcze w kontekście jurysdykcji gospodarczej oraz prawo bilansowe. Kwestie te zostały przeanalizowane w jej pracy na tle ogólnych dysfunkcji w gospodarce wynikających z modelu ustroju gospodarczego.

Demografią przedsiębiorstw z wykorzystaniem analizy historii zdarzeń zajmowała się Ptak-Chmielewska (2010). Autorka ta w analizach korzystała m.in. z modelu proporcjonalnego hazardu Coxa. W pracy Dehnela (2010) można znaleźć przede wszystkim podejście statystyczne do badania mikroprzedsiębiorczości oparte na przekrojach regionalno-branżowych.

Markowicz (2012) zajęła się systematyką metodyki analizy żywotności firm w Polsce. Przedmiotem zainteresowania autorki było połączenie analizy wskaźnikowej z metodami analizy czasu trwania przedsiębiorstw. Do analizy żywotności firm, badania wpływu na nią wybranych cech firmy oraz oceny ryzyka likwidacji zastosowała tablice przetrwania, estymator Kaplana-Meiera, model Coxa i model logitowy.

Z kolei opracowanie Pocięchy (2012) jest poświęcone wszechstronnej analizie zastosowań modelu logitowego, jego zalet i ograniczeń, w kontekście prognozowania bankructw.

Nehrebecka i Dzik (2013) przedstawiły model przypisujący prawdopodobieństwo bankructwa przedsiębiorstwa na podstawie indywidualnych danych bilansowych i raportu dochodów polskich spółek według danych za okres 2001—2010 zgromadzonych przez GUS. Autorki zbadały determinanty sygnałów ostrzegawczych oraz możliwości wczesnej identyfikacji tych sygnałów. W badaniu wykorzystano regresję logistyczną. Autorki podjęły także 3-letnią próbę prognozy bankructw analizowanych przedsiębiorstw.

W tym artykule dokonano analizy przeżywalności firm funkcjonujących w województwie małopolskim w głównej mierze na podstawie rocznych zestawień nowych i wyrejestrowanych podmiotów z lat 2006—2014. W badanym okresie (01.01.2006—31.08.2014 r.) zarejestrowanych w REGON było ok. 260 tys. podmiotów gospodarki narodowej i równolegle wyrejestrowano niecałe 200 tys. podmiotów. W każdym z rocznych zestawień znajdują się informacje na temat dokładnej daty rozpoczęcia działalności gospodarczej (ewentualnie daty zakończenia działalności) powiatu i gminy, w której podmiot był zarejestrowany, kodu PKD i wielkości firmy (mierzonej deklarowaną liczbą pracowników). Analizę zaprezentowaną w artykule oparto m.in. na modelach statystycznych, przeprowadzono także testy na istotność różnic w prawdopodobieństwie przeżycia firm według przynależności do poszczególnych grup.

MODELE PRZEŻYCIA PRZEDSIĘBIORSTWA

Przyjmuje się, że w badaniu uwzględniono N homogenicznych przedsiębiorstw, dla których czas od powstania do likwidacji jest notowany jako T_i , $i = 1, \dots, N$. Zmienne T_i są niezależnymi realizacjami zmiennej losowej T , która w szczególności przyjmuje wyłącznie wartości dodatnie. Rozkład zmiennej losowej T zdefiniowany jest za pomocą dystrybuanty prawdopodobieństwa $F(t) = P(T \leq t)$. Dystrybuanta jest jedną z pięciu podstawowych funkcji wykorzystywanych w analizie przeżycia. Obok niej są ważne następujące funkcje:

- gęstości prawdopodobieństwa, czyli opisująca prawdopodobieństwo likwidacji przedsiębiorstwa w chwili t :

$$f(t) = \frac{dF(t)}{dt} \quad (1)$$

- przeżycia (inaczej nazywana funkcją dożycia lub funkcją trwania), która opisuje prawdopodobieństwo, że podmiot nie zostanie zlikwidowany do chwili t :

$$S(t) = 1 - F(t) = P(T \geq t) \quad (2)$$

- hazardu (inaczej funkcja intensywności), czyli opisująca prawdopodobieństwo, że podmiot zostanie zlikwidowany w chwili t pod warunkiem, że do tej chwili nadal istnieje:

$$h(t) = \frac{f(t)}{S(t)} \quad (3)$$

- skumulowana funkcja hazardu:

$$H(t) = \int_0^t h(s) ds = -\ln S(t) \quad (4)$$

Wzór (4) prowadzi do konkluzji, że skumulowana funkcja hazardu mierzy całkowite ryzyko likwidacji skumulowane od momentu powstania do chwili t .

Warto uświadomić sobie, że to właśnie funkcja hazardu jest kluczem do zrozumienia nowoczesnej analizy przeżycia. W analizie przeżycia większą wagę przypisuje się do pary funkcji hazardu i skumulowanej funkcji hazardu niż do pary dystrybuanty i funkcji gęstości prawdopodobieństwa. W szczególności, znając postać zwykłej i skumulowanej funkcji hazardu, można otrzymać wspomniane wcześniej funkcje, wykorzystując zależności:

$$S(t) = \exp\{-H(t)\} \quad (5)$$

$$F(t) = 1 - \exp\{-H(t)\} \quad (6)$$

$$f(t) = h(t) \exp\{-H(t)\} \quad (7)$$

Badanie empiryczne oparte jest na danych, które opisują zjawisko w ustalonym okresie. Praca nad tego typu danymi wymaga uwzględnienia tzw. obserwacji cenzurowanych (lub inaczej — uciętych). Dostępne dane wymuszają zakończenie badania w z góry ustalonym momencie T_i^C , czyli okresie cenzurowania dla podmiotu o numerze i . W związku z tym nie jest możliwe zaobserwowanie wszystkich wartości T_i . Obserwuje się za to pary (Y_i, δ_i) , $i = 1, \dots, N$, gdzie:

$$Y_i = \begin{cases} T_i, & T_i \leq T_i^C \\ T_c, & T_i^C < T_i \end{cases} \quad \text{oraz} \quad \delta_i = \begin{cases} 1, & T_i \leq T_i^C \\ 0, & T_i^C < T_i \end{cases} \quad (8)$$

Zmienna Y_i zawiera informację o czasie trwania przyporządkowanym właściwemu podmiotowi (czas trwania lub czas od momentu powstania do momentu cenzurowania), natomiast δ_i ma wartość 1, gdy podmiot został zlikwidowany i wartość 0 w przeciwnym przypadku.

Nieparametryczne modele przeżycia to modele, w których nie określa się postaci analitycznej rozkładu zmiennej losowej T , czyli czasu trwania. Pomimo tego, że podejście takie jest uproszczone, to dostarcza ono samo w sobie wielu użytecznych informacji i często stosuje się je jako wstępne badanie przed właściwą analizą parametryczną.

Historycznie pierwszymi znanymi elementami nieparametrycznej analizy przeżycia są tablice trwania życia, nazywane też czasem tablicami wymieralności¹. W artykule wykorzystano tradycyjną metodę budowy tablic trwania, nazywaną również metodą aktuarialną (Chiang, 1984). W danych wykorzystanych do budowy tablic w tej pracy nie ma informacji dotyczących liczby podmiotów zawieszających działalność gospodarczą, dlatego standardowa notacja zostanie uproszczona.

W analizie przeżycia wykorzystującej tablice wymieralności zakłada się, że badany okres jest podzielony na równe przedziały czasu. Niech D_i oznacza liczbę podmiotów gospodarczych likwidowanych w okresie i , natomiast N_i — liczbę podmiotów, które były zarejestrowane na początku tego okresu. Prawdopodobieństwo likwidacji przedsiębiorstwa w okresie i , pod warunkiem że jest ono nadal zarejestrowane, szacowane jest za pomocą wzoru:

$$\hat{q}_i = \frac{D_i}{N_i} \quad (9)$$

Naturalną konsekwencją zależności (9) jest wzór na warunkowe prawdopodobieństwo przeżycia okresu i , który przyjmuje postać:

$$\hat{p}_i = 1 - \hat{q}_i = 1 - \frac{D_i}{N_i} \quad (10)$$

Aby oszacować wartości funkcji przeżycia $S(t)$, należy obliczyć odpowiedni iloczyn prawdopodobieństw warunkowych. Estymator funkcji przeżycia dany jest zatem wzorem:

¹ Najstarsze tablice trwania życia przypisywane są J. Gauntowi, który w 1662 r. w książce *Natural and Political Observations Made Upon the Bills of Mortality* stworzył takie tablice dla wymieralności mieszkańców Londynu. Najstarsze znane polskie tablice trwania życia zostały stworzone przez J. Słonimskiego i A.B. Danielewicza w XIX w. i dotyczyły długości życia mieszkańców Warszawy.

$$\hat{S}_{TAB}(t) = \prod_{i|t_i \leq t} \left(1 - \frac{D_i}{N_i}\right) \quad (11)$$

Otrzymana w ten sposób funkcja ma formę schodkową, czyli jest stała w poszczególnych okresach. Funkcje hazardu dla tablic trwania tworzonych na podstawie metody aktuarialnej otrzymuje się z zależności:

$$\hat{h}_{TAB}(i) = \frac{D_i}{\left(N_i - \frac{1}{2}D_i\right) \cdot |i|} \quad (12)$$

gdzie przez $|i|$ rozumiana jest długość jednostki czasu. W naszym przypadku liczba analizowanych zdarzeń jest stosunkowo duża. Zatem jest spełniony warunek, by przedstawione estymatory nie były obciążone (Landmesser, 2013).

Najczęściej wykorzystywanym nieparametrycznym estymatorem funkcji przeżycia jest jednak estymator Kaplana-Meiera postaci:

$$\hat{S}_{KM}(t) = \prod_{i|t_i \leq t} \left(1 - \frac{D_i}{N_i}\right) \quad (13)$$

gdzie N_i jest liczbą podmiotów gospodarczych zarejestrowanych na początku okresu t_i , natomiast D_i to liczba podmiotów zlikwidowanych w tym okresie. Definicja estymatora jest ładząco podobna do tej, która jest stosowana w przypadku konstrukcji tablic wymieralności metodą aktuarialną (11), jednakże jest tutaj pewna subtelna, ale zarazem istotna różnica. W tym przypadku nie ma założenia o podziale analizowanego okresu na odcinki o równej długości — każdy z okresów t_i traktuje się jako zmienną losową. Podejście takie jest usprawiedliwione tym, że w każdym okresie t_i ma miejsce co najmniej jedno zdarzenie.

Do obliczenia błędów standardowych estymatora Kaplana-Meiera wykorzystuje się formułę Greenwooda (Greenwood, 1926):

$$s(\hat{S}_{KM}(t)) = \hat{S}_{KM}(t) \sqrt{\sum_{i|t_i \leq t} \frac{D_i}{N_i(N_i - D_i)}} \quad (14)$$

Z kolei funkcja hazardu dana jest zależnością:

$$\hat{h}_{KM}(t_i) = \frac{D_i}{N_i(t_{i+1} - t_i)} \quad (15)$$

Estymator funkcji hazardu jest obarczony błędem standardowym postaci:

$$s(\hat{h}_{KM}(t_i)) = \frac{\hat{h}_{KM}(t_i)}{\sqrt{D_i}} \quad (16)$$

Przyjmuje się zwykle, że dla okresów wykraczających poza zakres badania wartość estymatora jest stała. Estymator Kaplana-Meiera wykazuje dużą zmienność w okresach, w których liczba obserwacji jest niewielka. W przypadku analizy prezentowanej w tym opracowaniu nie jest to problemem ze względu na obszerność zbioru danych.

Analiza nieparametryczna pozwala na porównanie funkcji przeżycia w poszczególnych podgrupach składających się na rozpatrywaną populację. Porównania te mogą być dokonywane na dwa zasadnicze sposoby. Pierwszy z nich opiera się na analizie graficznej. Dzięki wykresom łatwo bowiem zauważyć, czy i jakie różnice występują pomiędzy analizowanymi grupami. Drugi sposób oparty jest na zastosowaniu jednego z wielu dostępnych nieparametrycznych testów statystycznych, sprawdzających równość odpowiadających im funkcji przeżycia.

Najbardziej znanym nieparametrycznym testem na istotność różnic jest test Logrank, nazywany również testem Mantela-Haenszela (Mantel i Haenszel, 1959) lub uogólnionym testem Savage'a (Savage, 1956). Test może być przeprowadzony w ogólnej postaci dla dowolnej skończonej liczby grup. Najlepsze rezultaty otrzymać można w przypadku, gdy funkcje hazardu dla poszczególnych grup są wzajemnie proporcjonalne. Hipoteza zerowa zakłada równość funkcji przeżycia w podgrupach. Statystyka empiryczna Q w teście przyjmuje postać:

$$Q = \frac{(O - E)^2}{\text{Var}(O - E)} \quad (17)$$

gdzie $O - E$ stanowi różnicę pomiędzy liczbą zaobserwowanych i liczbą oczekiwanych zdarzeń, natomiast $\text{Var}(O - E)$ oznacza wariancję obliczoną dla tej różnicy². Przy założeniu, że hipoteza zerowa jest prawdziwa statystyka testowa ma rozkład χ^2 o $k - 1$ stopniach swobody (w przypadku testu dla k grup).

Różnice pomiędzy testem Logrank a innymi testami tego typu wynikają ze sposobu zdefiniowania wag. Wagi te odnoszą się do różnic pomiędzy wartościami funkcji przeżycia w podgrupach. W standardowym teście Logrank przyjmuje się, że wagi są jednakowe w całym badanym okresie — $W(t_i) = 1$. Oznacza to, że (przykładowo) różnice na początku badania są równie istotne, jak na końcu, bez względu na liczbę podmiotów, których zdarzenie dotyczy.

Z kolei w teście Gehana-Breslowa (Gehan, 1965; Breslow, 1970), nazywanym czasem testem Wilcozona, jako wagi przyjmuje się liczbę jednostek, które

² W szczególności wzór przyjmuje postać $O - E = \left(\sum_j W(t_j)(m_{ij} - e_{ij}) \right)^2$, gdzie $W(t_j)$ ozna-

cza wagę przypisaną dla okresu t_j , m_{ij} reprezentuje liczbę likwidacji w grupie i w okresie t_j , natomiast e_{ij} odpowiada oczekiwanej liczbie likwidacji w grupie i w okresie t_j , liczoną dla każdego okresu dzięki porównaniu proporcji podmiotów narażonych na ryzyko w grupie do łącznej liczby likwidacji.

w chwili t są narażone na ryzyko — $W(t_i) = N_i$. Kolejnym testem należącym do tej rodziny jest test Tarone-Ware (Tarone i Ware, 1977), w którym za wagi przyjmuje się pierwiastek ze wspomnianej liczby jednostek narażonych na ryzyko — $W(t_i) = \sqrt{N_i}$.

Test Gehana-Breslowa ma największą moc, gdy funkcja przeżycia pochodzi z rozkładu log-normalnego, natomiast charakteryzuje się tendencją do zbyt częstego odrzucania hipotezy zerowej w sytuacjach, gdy jest ona prawdziwa dla danych zawierających znaczną liczbę obserwacji cenzurowanych. Prentice (1978) pokazał, że w wielu przypadkach test Tarone-Ware jest mocniejszy zarówno od testu Logrank, jak i od testu Gehana-Breslowa.

Test Peto-Peto-Prentice (Peto R. i Peto J., 1972; Kalbfleisch i Prentice, 1980) jest testem, którego autorzy podeszli inaczej do problemu ustalenia wag niż poprzednicy. Zdefiniowali estymator funkcji przeżycia:

$$\hat{S}(t) = \prod_{i|t_i \leq t} \left(1 - \frac{D_i}{N_i + 1}\right) \quad (18)$$

a następnie jako wagi przyjęli zdefiniowane przez siebie estymatory — $W(t_i) = \hat{S}(t_i)$. Zaletą tego testu jest fakt, że nie traci on mocy w przypadku różnego cenzurowania w grupach.

Test Fleminga-Harringtona jest ostatnim testem należącym do tej grupy, który został wykorzystany w tym artykule. Autorzy testu (Fleming i Harrington, 1991) jako wagi przyjęli iloczyn funkcji przeżycia i dystrybuanty rozkładu ważony odpowiednimi potęgami:

$$W(t_i) = \left[\hat{S}(t_{i-1})\right]^p \left[1 - \hat{S}(t_{i-1})\right]^q \quad (19)$$

gdzie $\hat{S}(t_{i-1})$ jest estymatorem funkcji przeżycia w poprzednim okresie, natomiast p i q są parametrami, których specyfikacja może być dowolnie wybrana przez badacza. W przypadku gdy $p < q$, większa waga jest przypisywana podmiotom mającym dłuższe czasy trwania, natomiast dla $p > q$ jest odwrotnie.

Pomimo relatywnie dużej liczby testów nieparametrycznych służących porównaniu funkcji przeżycia, nie ma spójnej grupy kryteriów pozwalających rozstrzygnąć, który z testów ma największą moc i powinien być wykorzystany w analizach. Badania przeprowadzone w pracy Latta (1981) wskazują, że moc testów zmienia się w zależności od wielkości badanej próby, od mechanizmu cenzurowania i rozkładu gęstości funkcji hazardu.

CHARAKTERYSTYKA ZBIORU DANYCH I JEGO STRUKTURY

Dane wykorzystane w artykule pochodzą z Urzędu Statystycznego w Krakowie. Zbiór danych zawiera informacje dotyczące podmiotów gospodarki narodowej zarejestrowanych w Małopolsce. Informacje tam zawarte dotyczą wszystkich podmiotów rejestrowanych oraz wyrejestrowanych w latach 2006—2014. W szczególności za pomocą tej bazy danych możliwa jest analiza z podziałem na: wielkość zatrudnienia, sekcje PKD, powiaty, w których zarejestrowano podmiot gospodarki narodowej oraz rozróżnienie miejsca założenia (miasto/wieś).

Zbiór danych zawiera informacje na temat wszystkich podmiotów zarejestrowanych w REGON pomiędzy 1 stycznia 2006 r. a 31 sierpnia 2014 r. Łącznie było to 267755 podmiotów, z czego 263311 powstało w tym okresie. Pozostałe 5444 podmioty zostały zarejestrowane w latach 2006—2014, jednakże faktyczna data rozpoczęcia przez nie działalności przypadała przed badanym okresem. Najstarszym podmiotem gospodarki narodowej, któremu nadano numer REGON w tym okresie w Małopolsce, był podmiot założony w powiecie andrychowskim, dla którego faktyczną datą założenia był 20 stycznia 1796 r.³ Podmiot ten jest w grupie 0—9 zatrudnionych i został zakwalifikowany do sekcji S⁴ w PKD 2007. Fakt pojawiania się tego typu podmiotów w oficjalnej statystyce tłumaczyć można na dwa sposoby. Po pierwsze, mogą to być podmioty, które wcześniej nie miały obowiązku rejestrowania się w bazie⁵. Po drugie, mogą to być podmioty, które zakończyły uprzednią działalność, a później zaczęły inną⁶. W analizowanym okresie wyrejestrowano równocześnie 197022 podmioty gospodarki narodowej, z czego dla 91281 data założenia działalności gospodarczej przypadała na lata 2006—2014.

Dla każdego z zarejestrowanych podmiotów dostępna jest informacja na temat symbolu dominującej działalności gospodarczej obowiązującej w chwili powstania podmiotu (PKD 2004, PKD 2007 lub oba symbole jednocześnie⁷),

³ Na podstawie posiadanego zbioru danych autorzy nie są w stanie rozróżnić, czy wartość jest faktyczną datą rozpoczęcia działalności czy też wynikiem błędu podczas wprowadzania danych do systemu.

⁴ Podklasa 94.91.Z oznaczająca „Działalność organizacji religijnych”.

⁵ Przykładowo 01.08.2012 r. weszły w życie przepisy z 15.04.2011 r. o systemie informacji oświatowej, które nakładają obowiązek posiadania numeru REGON przez każdą istniejącą i nowo powstałą szkołę oraz placówkę oświatową w Polsce.

⁶ Numer REGON po zakończeniu działalności gospodarczej nie jest przekazywany innemu podmiotowi, w przypadku zakładania nowej działalności „wraca” do poprzedniego właściciela, dlatego jako data rozpoczęcia działalności figuruje data, która czasem znacznie poprzedza datę złożenia wniosku.

⁷ Aby poprawić czytelność uzyskanych wyników w przypadku każdego z podmiotów, dla którego dostępny był wyłącznie symbol działalności według PKD 2004, przekonwertowano ten symbol do jego odpowiednika z PKD 2007. W tym celu wykorzystano publikowany przez GUS klucz powiązań pomiędzy PKD 2004 i PKD 2007.

daty powstania podmiotu i zakończenia działalności, symbole powiatu i gminy (na podstawie tego symbolu możliwe jest odczytanie, czy podmiot gospodarki narodowej powstał w mieście czy na wsi) oraz deklarowana wielkość zatrudnienia, która należy do jednej z pięciu grup: 0—9, 10—49, 50—249, 250—999, 1000 i więcej pracowników⁸.

Na potrzeby badania empirycznego przyjęto schemat, według którego każdy z podmiotów powstał w chwili t_0 ⁹, a następnie był obserwowany do momentu likwidacji lub do ostatniego dnia badania. W związku z tym dane są cenzurowane prawostronnie, ale czas cenzurowania jest inny dla poszczególnych przedsiębiorstw i uzależniony od daty powstania.

Szczegółowa lista zmiennych wraz z opisem i możliwymi do przyjęcia wartościami znajduje się w poniższym zestawieniu.

ZESTAWIENIE ZMIENNYCH OBJAŚNIAJĄCYCH WYKORZYSTANYCH W ANALIZIE NIEPARAMETRYCZNEJ

Zmienne	Opis zmiennych (w nawiasach podano odsetek obserwacji, których dana wartość dotyczy)
<i>Wielkość</i>	1 — 0—9 pracowników — mikroprzedsiębiorstwa (98,676%) 2 — 10—49 pracowników — małe przedsiębiorstwa (1,231%) 3 — 50—249 pracowników — przedsiębiorstwa średnie (0,087%) 4 — 250—999 pracowników — przedsiębiorstwa duże (0,005%) 5 — od 1000 pracowników — przedsiębiorstwa duże (0,002%)
<i>Sekcja</i>	A — rolnictwo, leśnictwo, łowiectwo i rybactwo (0,7%) B — górnictwo i wydobywanie (0,1%) C — przetwórstwo przemysłowe (7,8%) D — wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną, gorącą wodę i powietrze do układów klimatyzacyjnych (0,2%) E — dostawa wody; gospodarowanie ściekami i odpadami oraz działalność związana z rekultywacją (0,3%) F — budownictwo (17,8%) G — handel hurtowy i detaliczny; naprawa pojazdów samochodowych, włączając motocykle (24,5%) H — transport i gospodarka magazynowa (5,4%) I — działalność związana z zakwaterowaniem i usługami gastronomicznymi (4,5%) J — informacja i komunikacja (3,8%) K — działalność finansowa i ubezpieczeniowa (3,8%) L — działalność związana z obsługą rynku nieruchomości (2,2%) M — działalność profesjonalna, naukowa i techniczna (9,0%) N — działalność w zakresie usług administrowania i działalność wspierająca (4,2%) O — administracja publiczna i obrona narodowa; obowiązkowe zabezpieczenia społeczne (0,04%) P — edukacja (3,5%) Q — opieka zdrowotna i pomoc społeczna (4,4%) R — działalność związana z kulturą, rozrywką i rekreacją (1,8%) S — pozostała działalność usługowa (6,1%) T — gospodarstwa domowe zatrudniające pracowników; gospodarstwa domowe produkujące wyroby i świadczące usługi na własne potrzeby (0,0007%) U — organizacje i zespoły eksterytorialne (0,002%)

⁸ Procentowy udział poszczególnych grup w badanej populacji opisany w analizie nieparametrycznej.

⁹ Alternatywą jest schemat, w którym za dzień powstania przyjmuje się nie dzień zerowy, ale faktyczny dzień powstania. Nie jest to jednak możliwe w tym przypadku, ponieważ prowadząc badanie autorzy nie dysponowali informacjami na temat wszystkich podmiotów działających w danym okresie, w związku z czym otrzymane rezultaty nie mogłyby być uznane za wiarygodne.

**ZESTAWIENIE ZMIENNYCH OBJAŚNIAJĄCYCH WYKORZYSTANYCH
W ANALIZIE NIEPARAMETRYCZNEJ (dok.)**

Zmienne	Opis zmiennych (w nawiasach podano odsetek obserwacji, których dana wartość dotyczy)	
<i>Powiat^a</i>	1 — bocheński (2,5%) 2 — brzeski (2,0%) 3 — chrzanowski (3,0%) 4 — dąbrowski (1,0%) 5 — gorlicki (2,8%) 6 — krakowski (7,5%) 7 — limanowski (3,4%) 8 — miechowski (1,0%) 9 — myślenicki (3,3%) 10 — nowosądecki (5,5%) 11 — nowotarski (5,1%)	12 — olkuski (2,8%) 13 — oświęcimski (3,6%) 14 — proszowicki (0,8%) 16 — tarnowski (3,9%) 17 — tatrzański (2,7%) 18 — wadowicki (3,9%) 19 — wielicki (3,6%) 61 — Kraków (33,1%) 62 — Nowy Sącz (3,1%) 63 — Tarnów (3,4%)
<i>Miasto</i>	1 — dla podmiotów zarejestrowanych w miastach (61,0%); 0 — dla niezarejestrowanych w miastach (39,0%)	

a Oznaczenia powiatów przyjęte zgodnie ze standardami opisanymi w publikacji *Wykaz identyfikatorów i nazw jednostek podziału terytorialnego kraju*, GUS, Warszawa, 2010.

Źródło: opracowanie własne na podstawie danych Urzędu Statystycznego w Krakowie.

Z zestawienia można odczytać, że ponad 98% podmiotów gospodarki narodowej spośród 263311, które powstały w Małopolsce w badanym okresie, to podmioty zaliczane do grupy mikroprzedsiębiorstw, a nieco ponad 1% powstałych podmiotów zalicza się do grupy małych przedsiębiorstw. Łatwo zauważyć, że wraz ze zwiększaniem się deklarowanej liczby pracowników drastycznie zmniejsza się procentowy udział w grupie tworzonych podmiotów. W szczególności w badanym okresie powstało dwanaście przedsiębiorstw o zatrudnieniu od 250 do 999 pracowników i tylko cztery przedsiębiorstwa deklarujące zatrudnienie przynajmniej tysiąca pracowników.

Analizując firmy zatrudniające od 250 do 999 pracowników warto zauważyć, że spośród dwunastu firm powstałych w badanym okresie tylko jedna została zlikwidowana. W roku 2006 w Małopolsce istniało 299 firm tej wielkości. Średnia długość istnienia firm o tej wielkości, zlikwidowanych w badanym okresie, to piętnaście lat i pięć miesięcy.

Spośród największych firm (zatrudniających 1000 lub więcej pracowników) żadna z czterech powstałych w badanym okresie nie została zlikwidowana. W 2006 r. było ich w województwie 56. W badanym okresie powstały cztery takie firmy, a siedem zostało zlikwidowanych.

Znacznie mniej wyraźny rozdźwięk pomiędzy procentowym udziałem poszczególnych grup w badaniu występuje ze względu na podział na sekcje PKD 2007. Najwięcej podmiotów gospodarki narodowej, które powstały w latach 2006—2014 (bo 24,5%), zaklasyfikowanych jest do sekcji G (handel hurtowy i detaliczny; naprawa pojazdów samochodowych, włączając motocykle). Najmniejszą grupę stanowiły te zakwalifikowane do sekcji T (gospodarstwa domowe zatrudniające pracowników; gospodarstwa domowe produkujące wyroby i świadczące usługi na własne potrzeby) i U (organizacje i zespoły eksteryto-

rialne). Podobnie jak w przypadku dużych przedsiębiorstw, również dla tych dwóch sekcji mogą pojawić się problemy w przeprowadzeniu analizy parametrycznej.

Patrząc na podział danych ze względu na powiat, w którym podmioty zarejestrowano, uwagę zwraca miasto Kraków, gdzie powstał co trzeci z badanych podmiotów. Następny w kolejności jest powiat krakowski, gdzie w przybliżeniu powstało pięć razy mniej nowych podmiotów niż w stolicy województwa. Najmniejsza grupa nowych podmiotów zlokalizowana była w powiecie proszowickim. Ma to związek z niewielką powierzchnią powiatu i niższą niż przeciętnie gęstością zaludnienia. W sąsiadujących powiatach dąbrowskim i miechowskim powierzchnia i gęstość zaludnienia były jednak porównywalne do obserwowanych w powiecie proszowickim, a mimo to powstało tam niemal dziesięciokrotnie więcej podmiotów. We wszystkich powiatach województwa powstało wystarczająco dużo nowych podmiotów, by każdy z nich potraktować oddzielnie w prowadzonych badaniach.

Podział ze względu na podmioty miejskie i wiejskie został również ujęty w analizie. Okazuje się, że w miastach małopolskich powstaje o połowę więcej podmiotów niż na wsi. W województwie jest 61 miast i 14570 innych miejscowości, a mimo to dwa na pięć nowych podmiotów rejestruje się w trzech największych miastach województwa — Krakowie, Nowym Sączu i Tarnowie.

EMPIRYCZNA ANALIZA NIEPARAMETRYCZNA CZASU TRWANIA PRZEDSIĘBIORSTW W MAŁOPOLSCE

Właściwą nieparametryczną analizę przeżycia otwiera analiza przeprowadzona z wykorzystaniem tablic czasu trwania przedstawionych w tabl. 1. Ze względu na mnogość obserwacji prezentuje ona informacje w ujęciu rocznym.

Interesujący jest fakt, że w badanej populacji już w pierwszym dniu istnienia zlikwidowano 143 podmioty, zaś w pierwszym miesiącu przestało istnieć 966 nowych podmiotów. Przypuszczalnie część z nich nawet nie zaczęła właściwej działalności gospodarczej. W pierwszym roku zlikwidowano i wyrejestrowano ok. 8% nowych podmiotów gospodarki narodowej, po dwóch latach dołączyło do nich kolejne 9% podmiotów. Tendencja wzrostowa utrzymywała się również w trzecim roku, by osiągnąć 29% wyrejestrowanych podmiotów w pierwszych trzech latach od założenia. W następnych latach bezwzględna różnica podawana w punktach procentowych zmalała, by po ośmiu latach i ośmiu miesiącach zarejestrowanych pozostawało ok. 45% początkowej liczby podmiotów. Mediana czasu trwania wynosiła ok. siedmiu lat.

**TABL. 1. TABLICE CZASU TRWANIA DLA PODMIOTÓW Z REJESTRU REGON
W LATACH 2006—2014**

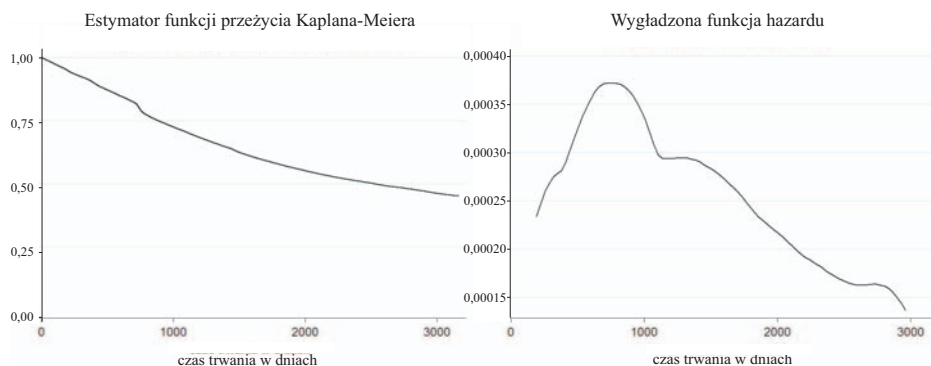
Czas trwania podmiotów	Funkcja przeżycia	Błąd standardowy	95% przedział ufności		Liczba		Dane cenzurowane ^a
					obserwacji na początku	podmiotów wyrejestrowanych	
1 rok	0,9193	0,0005	0,9183	0,9204	263311	20143	27223
2 lata	0,8253	0,0008	0,8237	0,8268	215945	20620	28818
3	0,7121	0,0010	0,7101	0,7140	166507	21183	24116
4	0,6409	0,0011	0,6387	0,6430	121208	11056	21253
5	0,5770	0,0012	0,5746	0,5794	88899	7777	21743
6	0,5320	0,0013	0,5294	0,5345	59379	4009	15985
7	0,4981	0,0014	0,4953	0,5009	39385	2059	14115
8 lat	0,4690	0,0016	0,4658	0,4722	23211	980	12878
96—104 miesiące	0,4461	0,0021	0,4220	0,4503	9353	234	9119

^a Dane cenzurowane oznaczają liczbę podmiotów, które po upływie wskazanego czasu nadal były zarejestrowane, ale ich dalsza obserwacja nie była możliwa.

Źródło: jak przy zestawieniu.

Analiza empirycznych wykresów funkcji przeżycia Kaplana-Meiera i wygładzonej funkcji hazardu przynosi dodatkowe ciekawe szczegóły. Wykr. 1 przedstawia wspomniane funkcje dla całej badanej populacji.

**Wykr. 1. FUNKCJA PRZEŻYCIA KAPLANA-MEIERA ORAZ WYKRES
WYGŁADZONEJ FUNKCJI HAZARDU**



Źródło: opracowanie własne na podstawie danych dziennych dla podmiotów gospodarki narodowej rejestrowanych w REGON w okresie 01.01.2006 — 31.08.2014 w Małopolsce.

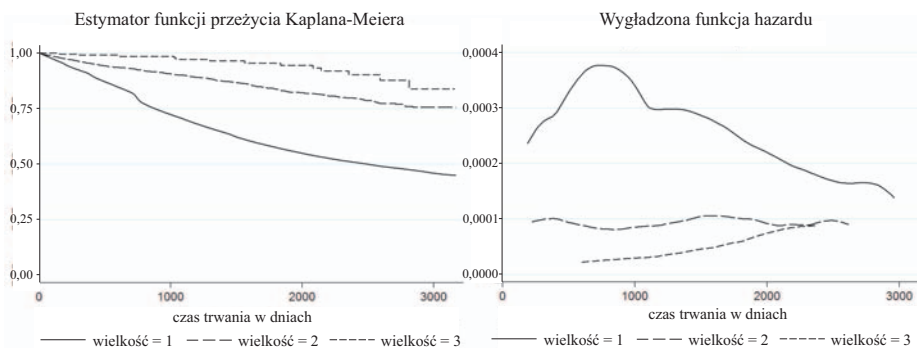
Wykres empirycznej funkcji przeżycia Kaplana-Meiera¹⁰ pokazuje, jaki procent populacji przetrwał opisaną na osi odciętych liczbę dni. Uwagę zwraca silny spadek wartości funkcji przeżycia w ok. 800 dniu istnienia i następujące po tym okresie zmniejszenie szybkości spadku wartości funkcji przeżycia. Wykres

¹⁰ Funkcja Kaplana-Meiera jest z natury funkcją schodkową, nie jest to jednak widoczne ze względu na obszerność zbioru danych.

obrazujący wygładzony przebieg empirycznej funkcji hazardu potwierdza te obserwacje. W trzecim roku istnienia podmioty narażone są na największy hazard, czyli należy brać pod uwagę warunkowe ryzyko likwidacji. Warunkiem jest oczywiście to, że firma nadal istnieje w tym okresie. Kształt funkcji hazardu zgadza się z teoretycznymi założeniami i wynikami empirycznymi innych badaczy (Agarwal, Sarkar i Echambadi, 2002; Cefis i Marsili, 2005; Zając, 2013). Funkcja hazardu przyjmuje kształt odwróconej litery U. Oznacza to, że ryzyko likwidacji w początkowym okresie istnienia firmy systematycznie wzrasta. Maksimum osiągnięte jest w trzecim roku istnienia, po czym następuje stopniowe zmniejszanie się hazardu. Z uwagi na występowanie cenzurowania prawostronnego w przypadku dużej liczby badanych podmiotów należy z dozą ostrożności interpretować gwałtowny spadek ryzyka likwidacji w ostatnich dniach badania. Efektów tego typu należy się spodziewać w badaniach empirycznych prowadzonych z wykorzystaniem modeli nieparametrycznych.

Jako pierwszy przeanalizowano podział ze względu na wielkość firmy, mierzona deklarowaną liczbą zatrudnionych pracowników (wykr. 2). Estymator Kaplana-Meiera, podobnie jak wygładzona funkcja hazardu dla mikroprzedsiębiorstw wyglądają niemal identycznie, jak odpowiednie wykresy dla całej populacji. Jest to naturalne w sytuacji, gdy ponad 98% podmiotów należy do tej grupy. Nowe informacje odczytać można z wykresów dotyczących małych i średnich przedsiębiorstw (grupy 2 i 3).

Wykr. 2. ESTYMATOR FUNKCJI PRZEŻYCIA KAPLANA-MEIERA ORAZ WYGŁADZONA FUNKCJA HAZARDU — PODZIAŁ ZE WZGLĘDU NA WIELKOŚĆ MIERZONĄ DEKLAROWANĄ LICZBĄ PRACOWNIKÓW



Źródło: opracowanie własne. Oznaczenia jak w tabl. 1.

Po ponad 3 tys. dni od dnia założenia w REGON figurowało nadal nieco ponad 75% małych przedsiębiorstw. Wykres empirycznej funkcji hazardu Kaplana-Meiera jest funkcją malejącą w sposób równomierny. Ma to odbicie w wygładzonej funkcji hazardu, która zachowuje się w sposób przypominający funk-

cję stałą. W przypadku przedsiębiorstw średnich można zaobserwować jeszcze inny wzorzec funkcji przeżycia i hazardu. Badany okres przetrwało niecałe 84% średnich firm. Funkcja przeżycia przez pierwsze 2 tys. dni spadała w sposób bardzo powolny. Po upływie tego czasu zauważalny był szybszy spadek wartości funkcji. Ze względu na zmniejszającą się wraz z upływem czasu liczbę obserwacji, poszczególne likwidacje wpływają na bardziej zauważalne schodki. Funkcja hazardu ma postać funkcji rosnącej. W opinii autorów, długość badanego okresu nie pozwala definitywnie rozstrzygnąć, czy funkcja hazardu jest monotoniczna. Być może od pewnego momentu wartość funkcji hazardu się stabilizowała, jednak aby móc zweryfikować tę hipotezę, należy przebadać przeżywalność firm w dłuższym okresie.

TABL. 2. TEST LOGRANK NA RÓWNOŚĆ FUNKCJI PRZEŻYCIA ZE WZGLĘDU NA ZMIENNĄ WIELKOŚĆ

Zmienna wielkość	Wartość	
	obserwowana	oczekiwana
1 (0— 9)	87641	86770,94
2 (10— 49)	406	1189,32
3 (50—249)	13	99,74
Suma	88060	88060,00
		$\chi^2 = 600,35^a$

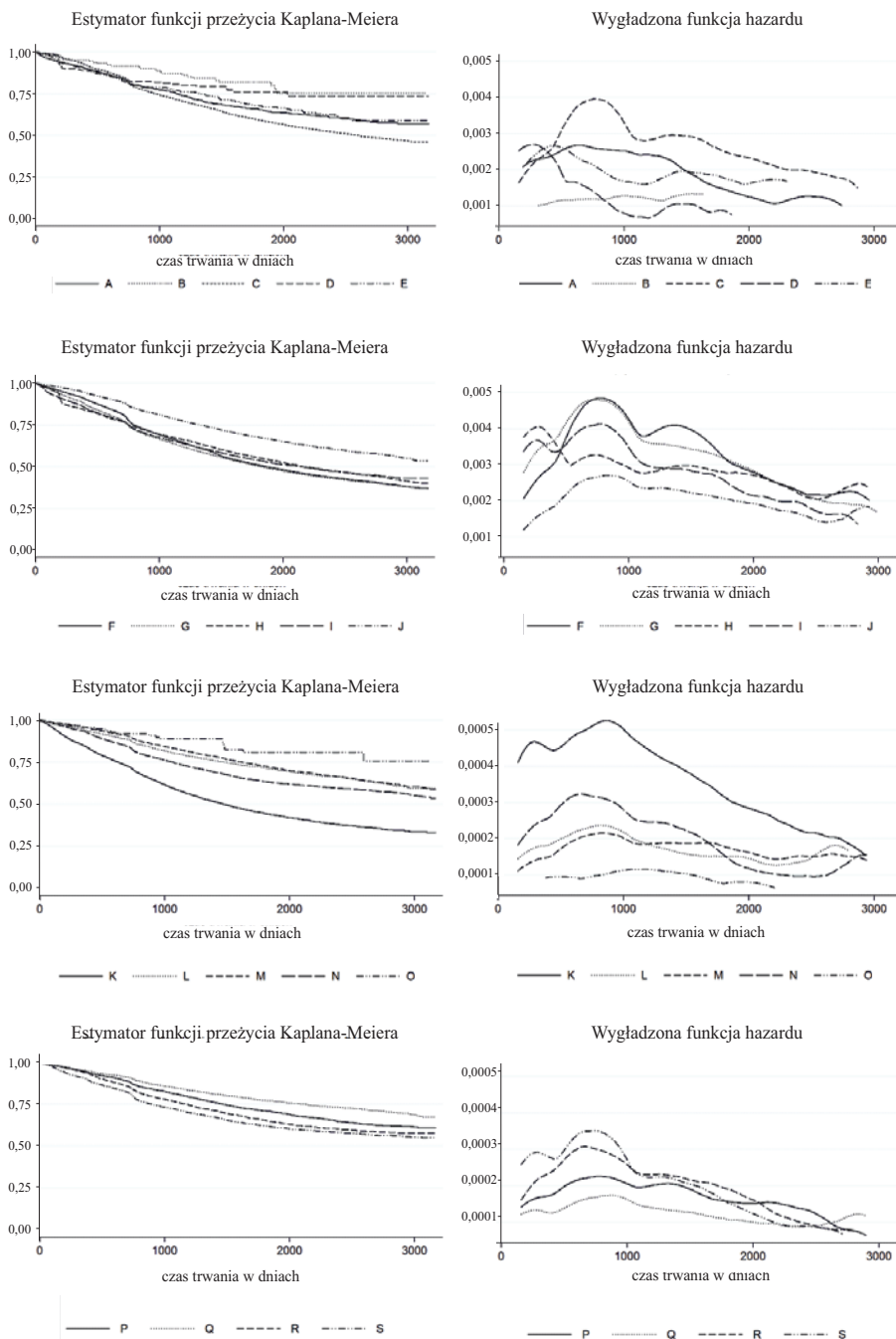
^a Oznacza, że poziom istotności statystyki wynosi 1%.
Z r ó d ł o: jak przy zestawieniu.

Aby potwierdzić statystyczną istotność podziału ze względu na wielkość firmy, wykonano testy nieparametryczne: Logrank, Gehan-Breslow, Tarone-Ware, Peto-Peto-Prentice i Fleming-Harrington. Wszystkie z tych testów mają tę samą hipotezę zerową mówiącą o równości w grupach. Każdy z nich odrzuca hipotezę zerową z prawdopodobieństwem przekraczającym 99%.

W związku z licznymi podobieństwami pomiędzy wspomnianymi testami i całkowitą zgodnością rezultatów prezentację wyników testu ograniczono do testu Logrank (tabl. 2).

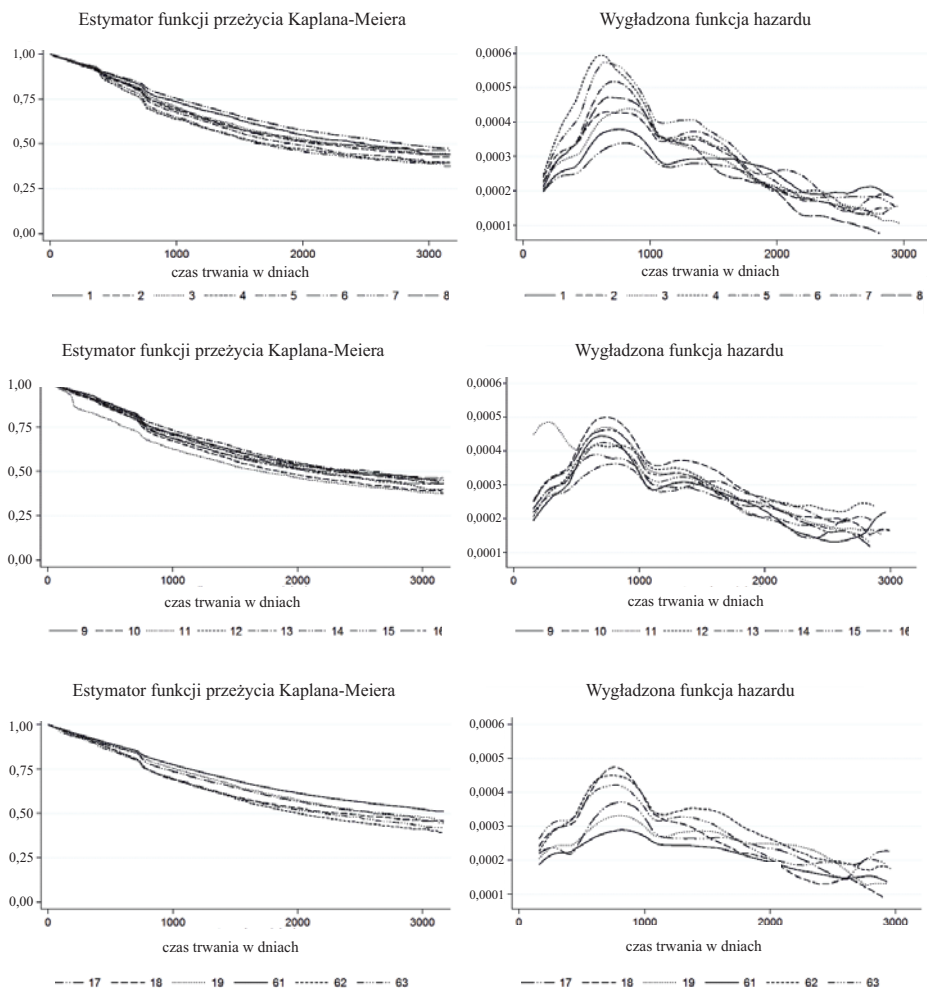
Analizując podział ze względu na przynależność do sekcji PKD 2007 łatwo zauważyć wyraźne rozróżnienie w ryzyku likwidacji w poszczególnych grupach (wykr. 3). Ekstremalne wartości przypadają dla sekcji K (działalność finansowa i ubezpieczeniowa) i sekcji O (administracja publiczna i obrona narodowa; obowiązkowe zabezpieczenia społeczne). Działalność finansowa i ubezpieczeniowa stanowiła branżę najbardziej narażoną na ryzyko likwidacji, a działalność administracyjna związana była z najmniejszym ryzykiem. Analizując wartości funkcji w sekcji K widać, że hazard był nawet pięciokrotnie większy niż w sekcji O. Badany okres obejmujący nieco ponad 3100 dni przetrwały trzy na cztery podmioty z sekcji O, równocześnie z sekcji K przetrwał mniej niż jeden na trzy założone podmioty. W największej pod względem liczby zakładanych podmiotów sekcji F ryzyko likwidacji jest podobne do tego z sekcji K.

Wykr. 3. ESTYMATOR FUNKCJI PRZEŻYCIA KAPLANA-MEIERA ORAZ WYGŁADZONA FUNKCJA HAZARDU — PODZIAŁ ZE WZGLĘDU NA PRZYNALEŻNOŚĆ DO SEKCJI PKD 2007



Źródło: jak przy wykr. 2.

Wykr. 4. ESTYMATOR FUNKCJI PRZEŻYCIA KAPLANA-MEIERA ORAZ WYGŁADZONA FUNKCJA HAZARDU — PODZIAŁ ZE WZGLĘDU NA POWIAT, W KTÓRYM ZAREJESTROWANO PODMIOT GOSPODARKI NARODOWEJ



Źródło: jak przy wykr. 2.

Wygładzona funkcja hazardu przybiera U-kształtną postać dla niemal wszystkich badanych sekcji, przy czym wyjątkiem jest sekcja B (górnictwo i wydobywanie), gdzie hazard ogólnie wydaje się rosnać, pomimo że podlega niewielkim wahaniom.

Nieparametryczne testy równości zgodnie wskazywały na istotne statystycznie różnice w ryzyku likwidacji ze względu na przynależność do sekcji PKD 2007 (tabl. 3).

**TABL. 3. TEST LOGRANK NA RÓWNOŚĆ FUNKCJI PRZEŻYCIA
ZE WZGLĘDU NA ZMIENNĄ SEKCJA**

Zmienna sekcja	Wartość		Zmienna sekcja	Wartość	
	obserwowana	oczekiwana		obserwowana	oczekiwana
A	498	630,13	K	4627	3051,59
B	27	66,06	L	1296	2050,28
C	6610	7070,39	M	4802	8293,33
D	70	109,54	N	3293	4064,46
E	183	244,57	O	18	50,86
F	18380	15559,21	P	2094	3304,32
G	26000	20634,23	Q	2029	4186,17
H	5224	4522,59	R	1308	1704,45
I	4427	3734,46	S	4808	5371,73
J	2366	3411,64			$\chi^2=7016,60^a$

^a Jak przy tabl. 2.

Źródło: jak przy zestawieniu.

Podział ze względu na powiaty, w których zarejestrowano działalność gospodarczą, również okazuje się mieć wpływ na długość trwania podmiotów (wykr. 4). Wyglądzona funkcja hazardu przyjmuje największe punktowe wartości dla powiatów dąbrowskiego i limanowskiego. W powiatach tych ponad 60% zarejestrowanych podmiotów zlikwidowano po upływie analizowanego okresu. Najmniejszy hazard obserwowano dla działalności gospodarczej prowadzonej w Krakowie — badany okres przetrwała połowa założonych podmiotów.

Postać funkcji hazardu wydaje się być podobna przy podziale ze względu na powiaty, jako wyróżniający się należy wskazać powiat nowotarski. W powiecie tym wyraźnie zauważalne było przesunięcie okresu największego ryzyka likwidacji na koniec pierwszego roku działalności i utrzymywanie się wysokiego hazardu przez kolejne dwa lata. Po upływie badanego okresu w powiecie nowotarskim nadal zarejestrowane w REGON było ok. 40% podmiotów (podobnie jak w powiatach dąbrowskim i limanowskim).

Jak już wspomniano, wśród miast na prawach powiatu historycznie najmniejsze ryzyko likwidacji wystąpiło w Krakowie. W Nowym Sączu i w Tarnowie ryzyko to było znacznie większe niż w Krakowie i znajdowało się powyżej średniej dla całego województwa.

Nieparametryczne testy równości po raz kolejny zgodnie wskazują na istotne statystycznie różnice w ryzyku likwidacji, tym razem ze względu na powiat, w którym zarejestrowano działalność gospodarczą (tabl. 4).

**TABL. 4. TEST LOGRANK NA RÓWNOŚĆ FUNKCJI PRZEŻYCIA
ZE WZGLĘDU NA ZMIENNĄ POWIAT**

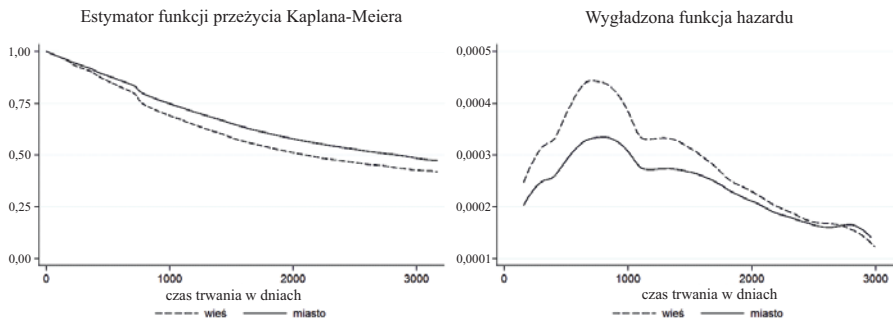
Zmienna powiat	Wartość		Zmienna powiat	Wartość	
	obserwowana	oczekiwana		obserwowana	oczekiwana
1	2164	2194,34	12	2739	2511,45
2	1906	1705,80	13	3173	3169,53
3	2823	2633,87	14	724	702,19
4	6146	6761,53	15	1763	1653,67
5	2889	2422,18	16	3733	3280,09
6	6146	6761,53	17	2386	2490,20
7	3603	2779,04	18	3617	3341,53
8	946	868,73	19	2929	3186,97
9	2926	2866,44	61	24849	30311,56
10	5554	4550,57	62	3191	2790,52
11	5571	3996,12	63	3355	3012,44
$\chi^2=2570,40^a$					

^a Jak przy tabl. 2.

Źródło: jak przy zestawieniu.

Na koniec został do omówienia podział ze względu na rozróżnienie pomiędzy podmiotami gospodarki narodowej zakładanymi w gminach miejskich i wiejskich (wykr. 5). Wykres funkcji przeżycia Kaplana-Meiera wskazuje, że ryzyko związane z koniecznością zakończenia działalności gospodarczej było większe, gdy działalność była zakładana w gminie wiejskiej. Funkcja hazardu dla obu grup przyjęła podobną postać. Różnice w wielkości hazardu w pierwszych trzech latach istnienia zwiększyły się na niekorzyść wsi, ale po upływie tego okresu różnice między nimi zaczęły być coraz mniejsze, by po ośmiu latach zrównać się.

**Wykr. 5. ESTYMATOR FUNKCJI PRZEŻYCIA KAPLANA-MEIERA ORAZ
WYGŁADZONA FUNKCJA HAZARDU — PODZIAŁ ZE WZGLĘDU
NA DZIAŁALNOŚĆ PROWADZONĄ W MIASTACH I NA WSI**



Źródło: jak przy wykry. 2.

Podobnie jak we wcześniejszych przypadkach nieparametryczne testy równości były zgodne w ocenie statystycznej istotności różnic w ryzyku likwidacji wynikających z podziału na podmioty zakładane w miastach i na wsiach (tabl. 5).

TABL. 5. TEST LOGRANK NA RÓWNOŚĆ FUNKCJI PRZEŻYCIA ZE WZGLĘDU NA ZMIENNĄ MIASTO

Zmienna <i>miasto</i>	Wartość	
	obserwowana	oczekiwana
0	37209	32857,46
1	50851	55202,54
Suma	88060	88060,00
		$\chi^2=920,11^a$

a Jak przy tabl. 2.

Z r ó ł o: jak przy zestawieniu.

Wnioski

Badania empiryczne potwierdziły, że funkcja hazardu dla ryzyka likwidacji przedsiębiorstw w województwie małopolskim ma kształt odwróconej litery U. Ryzyko usunięcia z REGON na początku rośnie wraz z wiekiem firmy, swoje maksimum osiąga w 2—3 roku działalności, by po upływie tego okresu stopniowo się zmniejszać. Po upływie ośmiu lat od daty powstania nadal zarejestrowanych było w REGON ponad 45% podmiotów gospodarki narodowej. Mediana dla czasu trwania wynosiła ok. siedmiu lat.

Zarówno funkcje przeżycia, jak i hazardu dla przedsiębiorstw małopolskich różniły się w sposób statystycznie istotny ze względu na zróżnicowanie wielkości firm mierzonych deklarowaną liczbą pracowników, przynależnością do sekcji PKD 2007, powiatem, w którym prowadziły działalność gospodarczą i rozróżnieniem, czy firma działała w mieście czy też na wsi. Badania empiryczne potwierdziły, że podmioty gospodarki narodowej o większej liczbie pracowników istniały w przeciętnym ujęciu dłużej na rynku. Branżą najbardziej narażoną na ryzyko likwidacji przedsiębiorstwa była działalność finansowa i ubezpieczeniowa, a najmniej — działalność administracyjna. Działalność przedsiębiorstw trwała średnio najdłużej w Krakowie, a także w powiatach: krakowskim, wielickim i tatrzańskim. Statystycznie najkrócej istniały firmy zakładane w powiecie nowotarskim. Ponadto firmy zakładane w miastach działały średnio dłużej od firm zakładanych na wsiach.

Przeprowadzona analiza przeżycia obok rozlicznych zalet ma też ograniczenia. Wynikają one z braku uwzględnienia wpływu wskaźników makroekonomicznych na przedsiębiorstwa oraz skupieniu się na czynnikach, które oddziałują na istniejące już firmy.

prof. dr hab. Henryk Gurgul, dr Paweł Zając — AGH w Krakowie

LITERATURA

Agarwal, R., Audretsch, D.B. (2001). Does Entry Size Matter?: The Impact of the Life Cycle and Technology on Firm Survival. *Journal of Industrial Economics*, vol. 49. Blackwell Publishing.

Agarwal, R., Sarkar, M.B., Echambadi, R. (2002). The conditioning effect of time on firm survival: An industry life cycle approach. *Academy of Management Journal*, vol. 45, no. 5.

- Audretsch, D.B., Mahmoon, T. (1995). New Firm Survival: New Results Using a Hazard Function. *Review of Economics and Statistics*, vol. 77, no. 1.
- Breslow, N.E. (1970). A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship. *Biometrika*, vol. 57.
- Cefis, E., Marsili, O. (2005). A matter of life and death: Innovation and firm survival. *Industrial and Corporate Change*, vol. 14, no. 6.
- Chiang, C.L. (1984). *The Life table and its Applications*. Krieger Publ. Company, Malabar.
- Dehnel, G. (2010). *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*. Wydawnictwo Uniwersytetu Ekonomicznego. Poznań.
- Fleming, T.R., Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- Gehan, E.A. (1965). A Generalized Wilcoxon test for Comparing Arbitrarily Singly Censored Data. *Biometrika*, vol. 52.
- Gerosky, P.A. (1995). What do we Know About Entry. *International Journal of Industrial Organization*, vol. 13, no. 4.
- Gilbert, B.A., McDougall, P.P., Audretsch, D.B. (2006). New Venture Growth: A review and extension. *Journal of Management*, vol. 32.
- Gurgul, H., Zajac, P., Matschke, X., Matschke, M.J. (2014). *A Dynamic Model of Birth and Death of Enterprises in Germany*. *Betriebswirtschaftliche Forschung und Praxis*, vol. 66, no. 1.
- Hol, S. (2007). The influence of business cycle on bankruptcy probability. *International Transactions in Operational Research*, vol. 14, no. 1.
- Holmes, P., Hunt, A.J., Stone, I. (2009). An analysis of new firm survival using a hazard function. *Applied Economics*, vol. 42, no. 2.
- Kalbfleisch, J.D., Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Latta, R.B. (1981). A Monte Carlo study of some two-sample rank tests with censored data. *JASA*, vol. 76, no. 375.
- Landmesser, J.M. (2013). *Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce*. Wydawnictwo SGGW, Warszawa.
- Lin, P., Huang, D. (2008). Technological Regimes and Firm Survival: Evidence Across Sectors and Over Time. *Small Business Economics*, vol. 30.
- López-García, P., Puente, S. (2006). Business Demography in Spain: Determinants of Firm Survival. *Documentos de Trabajo*, vol. 608.
- Markowicz, I. (2012). Statystyczna analiza żywotności firm. *Rozprawy i Studia*, no. 835. Wydawnictwo Uniwersytetu Szczecińskiego.
- Mantel, N., Haenszel, W. (1959). Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. *Journal of the National Cancer Institute*, vol. 22.
- Mata, J., Portugal, P., Guimaraes, P. (1995). The Survival of New Plants: Start-up Conditions and Post-entry Evolution. *International Journal of Industrial Organization*, vol. 13, no. 4.
- Mączczyńska, E. (2009). Upadłości przedsiębiorstw — dysfunkcje, ich przyczyny. *Zeszyty Naukowe Małopolskiej Wyższej Szkoły Ekonomicznej w Tarnowie*, vol. 13, no. 2.
- Nehrebecka, N., Dzik, A.M. (2013). Business Demography in Poland: Microeconomic and Macroeconomic Determinants of Firm Survival. *Working Papers 2013-08, Faculty of Economic Sciences*. University of Warsaw.
- Nunes, A., Sarmiento, E. (2010). Business Demography Dynamics in Portugal: A Non-Parametric Survival Analysis. *GEE Papers 0022*, Gabinete de Estratégia e Estudos, Ministério da Economia e da Inovação, Lisbon.
- Peto, R., Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, Series A*, vol. 135, no. 2.
- Pociecha, J. (2012). Model logitowy jako narzędzie prognozowania bankructwa. Jego zalety i ograniczenia. W: *Spotkania z królową nauk*. Wydawnictwo Uniwersytetu Ekonomicznego, Kraków.

- Poznańska, K. (2005). Koncepcje przetrwania przedsiębiorstw i ich weryfikacja empiryczna w gospodarce polskiej. W: Mączyńska, E. *Ekonomiczne aspekty upadłości przedsiębiorstw w Polsce*. SGH, Warszawa.
- Prentice, R.L. (1978). Linear Rank Tests with Right-Censored Data. *Biometrika*, vol. 65.
- Ptak-Chmielewska, A. (2010). Analiza przeżycia przedsiębiorstw w Polsce na przykładzie wybranego województwa. W: Dittmann, P. (red.), *Prognozowanie w zarządzaniu firmą*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Wrocław.
- Savage, I.R. (1956). Contributions to the Theory of Rank-Order Statistics — The Two-Sample Case. *Annals of Mathematical Statistics*, vol. 27.
- Schumpeter, J.A. (1942). *Capitalism, Socialism and Democracy*. Harper and Brothers, New York.
- Śmiech, S. (2011). Analiza przeżycia podmiotów gospodarczych w województwie małopolskim w latach 2002—2008. *Zeszyty Naukowe UEK w Krakowie*, vol. 876.
- Tarone, R.E., Ware, J.H. (1977). On Distribution-Free Tests for Equality of Survival Distributions. *Biometrika*, vol. 77.
- Zając, P. (2013). The new approach to estimation of the hazard function in business demography on example of data from New Zealand. *Managerial Economics*, vol. 13.

Summary. *The aim of the article is to assess the survival time of companies operating in Małopolska voivodship in the years 2006—2014. The method of duration analysis was used in the study.*

The research results confirmed that the units of the national economy with a larger number of employees operated longer than the smaller enterprises. However, the most frequent liquidations involved companies from the finance and insurance, and the least — from administrative sphere. Concerning location aspect, the longest existence of the enterprises was observed in Kraków City and rural districts: Kraków, Wieliczka and Tatra. The shortest existing companies were noted in Nowy Targ powiat. The companies established in cities were active, on average, longer than the ones in villages.

Keywords: enterprises, Małopolska, hazard function, duration analysis.

Резюме. *Целью статьи является оценка выживания предприятий действующих в малопольском воеводстве в 2006—2014 гг. В обследовании был использован метод анализа выживания.*

Результаты обследования подтвердили, что субъекты национальной экономики с большим числом сотрудников дольше действовали, в то время ликвидация чаще всего касалась предприятий из сферы финансов и страхования, а реже всего это явление касалось административной сферы. В пространственном подходе деятельность предприятий продолжалась в среднем самый длинный период в Кракове и в повятах: краковском, велицком и в районе Татров. Короче всего выживали предприятия в новотарском повяте, а организованы в городах действовали в среднем дольше чем в деревнях.

Ключевые слова: предприятия, Малопольша, функция риска, анализ выживания.